

Multi-Criteria Ranking of Corporate Distress Prediction Models: Empirical Evaluation and Methodological Contributions

Abstract: Although many modelling and prediction frameworks for corporate bankruptcy and distress have been proposed, the relative performance evaluation of prediction models is criticised due to the assessment exercise using a single measure of one criterion at a time, which leads to reporting conflicting results. Mousavi et al. (2015) proposed an orientation-free super-efficiency DEA-based framework to overcome this methodological issue. However, within a super-efficiency DEA framework, the reference benchmark changes from one prediction model evaluation to another, which in some contexts might be viewed as “unfair” benchmarking. In this paper, we overcome this issue by proposing a slacks-based context-dependent DEA (SBM-CDEA) framework to evaluate competing distress prediction models. In addition, we propose a hybrid Cross-Benchmarking-Cross-Efficiency (CBCE) framework as an alternative methodology for ranking DMUs that are heterogeneous. Furthermore, using data on UK firms listed on London Stock Exchange (LSE), we perform a comprehensive comparative analysis of the most popular corporate distress prediction models; namely, statistical models, under both mono criterion and multiple criteria frameworks considering several performance measures. Also, we propose new statistical models using macroeconomic indicators as drivers of distress.

Keywords: Corporate Distress Prediction; Performance Criteria; Performance Measures; Context-Dependent Data Envelopment Analysis; Slacks-Based Measure

1. Introduction

Corporate credit and default risk drive important decisions in banking and finance. According to the Basel Committee on Banking Supervision (BCBS), default in credit risk refers to a failure of a borrower or counterparty to meet its obligations in accordance with agreed terms (Basel Committee on Banking Supervision, 2000, p. 1). Corporate credit studies have used the same parametric and non-parametric frameworks to predict business failure events such as credit default (e.g., Beaver, 1996), bankruptcy (e.g., Barboza et al., 2017; Ouenniche and Tone, 2017; Liang et al., 2016, Kim et al., 2016, Bauer and Agarwal, 2014; Tinoco and Wilson, 2013; Zhou, 2013; Hillegeist et al., 2004; Shumway, 2001; Wilson and Sharda, 1994; Ohlson, 1980), financial distress (e.g., Altman et al., 2017; Li et al., 2014, 2017; Sun et al., 2017; Zhou et al., 2015; Geng and Chen, 2015; Campbell et al., 2008; Bandyopadhyay, 2006), insolvency (e.g., Callejón et al., 2013; Jackson and Wood, 2013), and loan default (e.g., Jiang et al., 2017; Kou et al., 2014; Bhimani and Gulamhussen, 2013). Amongst the above-mentioned failure events, bankruptcy and distress events have been the subject of many prediction studies. Financial distress refers to a situation where a company cannot generate enough cash flows to fulfil its contractual obligations (Piesse et al., 2006, p. 478). Remaining in distress for a long time not only could impact adversely on the value of the company and the wealth of stockholders but also causes more financial and operational inefficiencies and could lead to ceasing the operation of the firm or bankruptcy. In general, Corporate bankruptcy causes significant losses to both the business community and the society as a whole - for details about the costs of bankruptcy, we refer the reader to Davydenko et al. (2012), Elkamhi et al. (2012), Branch (2002) and Gruber and Warner (1997). Therefore, early detection of a company's deteriorating condition or distress has such economic advantages that motivated both academics and practitioners to develop a range of corporate distress prediction models.

From a statistical point of view, a failure prediction model (FPM) is a typical classification problem, which uses the selected features; say accounting, market, and macroeconomic-based information, to classify firms into distress or non-distress categories or classes. During the last decades, numerous studies have employed different types of techniques from statistics, operational research (e.g., Li et al., 2017; Ouenniche and Tone, 2017; Avkiran and Cai, 2014; Li et al., 2014; Premachandra et al., 2011; Yeh et al., 2010; Premachandra et al., 2009), and artificial intelligence (e.g., Chen et al., 2016; Fethi and Pasiouras, 2010; Bahrammirzaee, 2010;

Charalambous et al., 2000) fields to design new failure prediction models. Initial studies on failure prediction use statistical techniques such as univariate discriminant analysis (e.g., Beaver, 1966, 1968), and multivariate discriminant analysis (e.g., Altman, 1968, 1973, 1983) as classification techniques. Later, conditional probability models such as linear probability models (e.g., Meyer and Pifer, 1970; Maddala, 1986), logit models (e.g., Martin, 1977; Ohlson, 1980) and probit models (e.g., Zmijewski, 1984) were used to predict the probability of failure. The common characteristic of these models, however, is that they are time-independent (i.e., static) in nature and as such fail to explicitly take time-varying features of a firm into account. Dynamic models such as survival (hazard) models (e.g., Lane et al., 1986; Crapp and Stevenson, 1987; Luoma and Laitinen, 1991; Shumway, 2001; Bharath and Shumway, 2008; Chava and Jarrow, 2004), and contingent claims models (e.g., Bharath and Shumway, 2008; Hillegeist et al., 2004) are the next group of models, which by design could take account of changes firms are facing or experiencing over time. Statistical techniques, however, are constrained by the potential severity of the underlying assumptions, i.e., linearity, multivariate normality, independence among predictor or input variables, and equal within-group variance-covariate matrices. Artificial intelligence and mathematical programming techniques are alternatives that overcome the methodological restrictions related to statistical techniques.

Considering the massive increase in the number of failure prediction models, a stream of the literature has focused on answering the question: which of these models are superior in performance? According to Zhou (2013), failure prediction models are data-fitting based empirical research consisting of a series of processes including sampling, features selection, modelling, and performance evaluation. Obviously, the performance of models is not only dependent on the sample selection, modelling techniques and feature selection procedures but also reliant on the evaluation process and the chosen performance criteria. In practice, several studies have compared the performance of competing models taking into account different modelling frameworks – e.g., Bauer and Agarwal (2014), Mousavi et al. (2015) and Wu et al. (2010); alternative sampling techniques – e.g., Neves and Vieira (2006), and Zhou (2013), and various features – e.g., Tinoco and Wilson (2013), Trujillo-Ponce et al. (2014). Furthermore, several criteria, including, *discriminatory power*, *calibration accuracy*, *information content* and *correctness of categorical prediction* have been used for the performance evaluation of alternative models.

Our survey of the existing studies concerned with the comparison of competing statistical prediction models, supports Bauer and Agarwal (2014) and Mousavi et al. (2015) arguments in addressing three main drawbacks in the related literature. Firstly, most of the existing studies failed to have a comprehensive comparison between all types of statistical prediction models, i.e. traditional statistical models, Contingent claim analysis (CCA) models and survival analysis models. Secondly, the existing literature has used a restricted number of criteria to evaluate the performance of competing models. Thirdly, as mentioned by Mousavi et al. (2015), the nature of the performance evaluation of competing prediction models remains mono-criterion, as they use a single measure of a single criterion at a time. Therefore, under mono-criterion evaluation, the rankings corresponding to different criteria are mostly different, which lead to a situation where practitioners cannot make a well-informed decision as to which model performs best when taking all criteria into account (e.g., Theodossiou, 1991; Bandyopadhyay, 2006; Tinoco and Wilson, 2013). To overcome this methodological drawback, Mousavi et al. (2015) proposed a multi-criteria assessment framework; namely, an orientation-free super-efficiency data envelopment analysis. However, within a super-efficiency DEA framework, the reference benchmark changes from one efficient DMU evaluation to another, which in some contexts might be viewed as “unfair” benchmarking (Ouenniche et al., 2014). In this study, we overcome this issue by proposing a variant of the context-dependent DEA (CDEA) framework proposed by Seiford and Zhu (2003) which embed SBM models in the layering procedure. To be more specific, we propose a slacks-based context-dependent DEA (SBM-CDEA) framework for evaluating the relative performance of competing distress prediction models. We use the proposed SBM-CDEA framework to compare the relative performance of the most commonly used and cited statistical corporate failure prediction modelling frameworks. In addition, to reduce the degree of heterogeneity between static and dynamic distress prediction models, we implement them within a rolling horizon framework. Furthermore, we propose a new methodology as an alternative for ranking prediction models, which takes account of the heterogeneous nature of prediction models; namely, a hybrid Cross-Benchmarking-Cross-Efficiency (CBCE) framework.

We organised models into three categories; namely, original models, refitted models, and new models. Last, but not least, we use different measures of four commonly used criteria in the

literature; namely, calibration accuracy, information content, the correctness of categorical prediction, and discriminatory power, to evaluate the relative performance of models.

The remainder of this paper is organised as follows. Section 2 provides a review of comparative studies related to competing statistical models. Section 3 explains the proposed context-dependent DEA framework as a multi-criteria methodology to compare the relative performance of competing distress prediction models. Section 4 explains the proposed Cross-Benchmarking-Cross-Efficiency framework for assessing corporate distress prediction models. Section 5 presents the research methodology. Then, section 6 presents the empirical results and discussions. Finally, section 7 outlines the main conclusions of the paper.

2. Existing Literature on Comparison of Competing Statistical Prediction Models

Since the existing literature on the comparative performance of failure prediction models is substantial, this section provides a review of the studies, which focus on comparisons of different types of statistical models; i.e., traditional statistical models, contingent claim analysis (CCA) models, and survival analysis (SA) models.

Panel I of Table 1 presents the comparison between traditional statistical models. From the introduction of univariate discriminant analysis by Beaver (1966) through the early years of the 1980s, the multivariate discriminant analysis (MDA) was the superior method for predicting corporate failure. From the 1980s until 2001, the logit (introduced by Ohlson, 1980) and probit (introduced by Zmijewski, 1984) models dominated statistical techniques.

Panel II of Table 1 presents the comparison between traditional statistical models and SA model. Shumway (2001) proposed the breakthrough discrete-time hazard (DTH) model – using a multi-period logit framework – for failure prediction. In theory, SA models take advantage of their dynamic structure, and therefore outperform traditional statistical models, which are static in nature. However, in practice, the results of comparative studies indicate that the type of information that models fed with have a significant impact on the performance of models and could overcome the design shortcomings of static models (Shumway, 2001); therefore, static models should not be discarded entirely.

Panel III of Table 1 presents the comparison between statistical models and CCA models. Hillgeist et al. (2004) proposed a Black-Scholes-Merton (BSM) based model that performs

better than two types of traditional statistical techniques; namely, logit and MDA. Reisz and Perlich (2007) compared the performance of three CCA models; namely, BSM, KMV, and Down-and-Out Call option (DOC) based models. Further, Agarwal and Taffler (2008) compared the performance of two types of market-based models; namely, Hillegeist et al. (2004) and Bharath and Shumway (2008) and the MDA model of Taffler (1984). The comparison results indicate that CCA models outperform traditional statistical models under most measures of performance.

Panel IV of Table 1 shows the comparison between CCA and hazard models. Campbell et al. (2008) compared the performance of a CCA model; namely, KMV (Kealhofer, McQuown and Vasicek) and two types of hazard models; namely, Shumway (2001) and Campbell et al. (2008). The results indicate that their suggested hazard model outperforms both KMV and Shumway (2001) models.

Panel V of Table 1 presents the comparison between CCA, hazard and traditional statistical models. Wu et al. (2010) compared the performance of three frameworks of traditional statistical models; namely, MDA model of Altman (1968), logit model of Ohlson (1980), probit model of Zmijewski (1984) with DTH model of Shumway (2001) and BSM-based model of Hillegeist et al. (2004). Bauer and Agarwal (2004) compared the performance of traditional statistical, CCA and DTH models. The results of both studies suggest that DTH model outperforms other models. However, there are conflicts in the ranking of other models regarding different measures.

[Insert *Table 1* Here]

In the next section, we shall present our methodological choices.

3. A Slack-based CDEA Framework for Assessing Corporate Distress Predictions

In this research, we propose an orientation-free non-radial (slacks-based measure) context-dependent DEA (SBM-CDEA) framework for evaluating the relative performance of competing corporate distress prediction models. Hereafter, we first present the SBM-CDEA framework. Then, we discuss how one might adapt it to evaluate the relative performance of competing corporate distress prediction models.

Data envelopment analysis (DEA), proposed by Charnes, Cooper and Rhodes (1978), is a linear programming techniques to assess the relative efficiency of a set of similar decision making units

(DMUs), where each DMU is considered as a system, which uses multiple inputs to produce a number of outputs. The decision variables of these linear programming models are the weights allocated to inputs and outputs, and these models are referred to as multiplier models. The objective function value of the chosen DEA model – commonly referred to as a DEA score, allows one to classify a DMU as being efficient or not depending on whether its DEA score is equal to 1 or not. In DEA terminology, the set of efficient DMUs is referred to as the efficient frontier and represents the empirical standard of excellence against which benchmarking is done. Solutions to DEA models allow one to identify the reference set or peer group to use for benchmarking each DMU in seeking improvements. For detailed presentations of different DEA models, the reader is referred to Cooper et al. (2006).

Following the pioneering study by Mousavi et al. (2015) in using non-radial (slacks-based measure), non-oriented super-efficiency DEA to evaluate the performance of bankruptcy prediction models, we propose the non-radial (slacks-based measure), non-oriented context dependent DEA framework as a device for multi-criteria ranking of distress prediction models. We use an orientation-free evaluation because we intend to assess distress prediction models and thus the choice between input-oriented or output-oriented analysis is irrelevant. Further, input-oriented and output-oriented DEA studies may result in different scores and rankings of DMUs. On the other hand, we use a non-radial framework because the radial DEA models may be infeasible for some DMUs, which could result in having ties in rankings. Furthermore, radial DEA models do not take account of possible excesses and shortfalls; namely, slacks, in inputs and outputs, respectively, which could result in over-estimating the efficiency scores due to ignoring mix efficiency. Finally, the reason to use context-dependent rather than super-efficiency scores to rank DMUs is that within the latter one, the scores are used to rank order the efficient DMUs; however, the efficient DMUs have different reference sets, which in some contexts could be considered as “unfair” benchmarking. On the other hand, within CDEA, a set of DMUs can be divided into different levels of efficient frontiers (evaluation context), and the attractiveness measure or the progress measure are used to rank those efficient DMUs belonging to the same specific evaluation context; that is, having the same level of efficiency or score. The proposed SBM-CDEA framework is summarised in the following stages:

Stage 1 – Returns-to-Scale (RTS) Analysis: Perform RTS analysis to find out which type of RTS to include in DEA models; that is, constant-returns-to-scale (CRS), increasing returns-to-

scale (IRS), decreasing return-to-scale (DRS), or variable returns-to-scale (VRS). There are many approaches to perform RTS analysis (e.g., Banker, 1984; Banker, Charnes and Cooper, 1984; Fare, Grosskopf and Lovell, 1994; Banker, Cooper, Thrall and Zhu, 2004; Tone, 2001a). These approaches are based on BCC models whether in multiplier form or envelopment form. In this paper, we use an approach based on SBM model for a better compatibility with the next stages of CDEA analysis. This approach could be summarised as follows:

Step 1: For each DMU_k ($k = 1, \dots, n$), solve the following SBM model of Tone (2001b) under CRS (i.e., without any additional constraint), under IRS (i.e., by augmenting the SBM model with the constraint $\sum_{j=1}^n \lambda_j \geq 1$), under DRS (i.e., by augmenting the SBM model with the constraint $\sum_{j=1}^n \lambda_j \leq 1$), and under VRS (i.e., by augmenting the SBM model with the constraint $\sum_{j=1}^n \lambda_j = 1$):

$$\begin{aligned}
\text{Min} \quad & \rho_k = \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{s_{i,k}^-}{x_{i,k}}\right) / \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_{r,k}^+}{y_{r,k}}\right) \\
\text{s.t.} \quad & \sum_{j=1}^n \lambda_j x_{i,j} + s_{i,k}^- = x_{i,k}; \forall i \\
& \sum_{j=1}^n \lambda_j y_{r,j} - s_{r,k}^+ = y_{r,k}; \forall r \\
& \lambda_j \geq 0; \forall j; s_{i,k}^- \geq 0, \forall i; s_{r,k}^+ \geq 0, \forall r
\end{aligned}$$

where the $x_{i,j}$ ($i = 1, \dots, m$) and $y_{r,j}$ ($r = 1, \dots, s$) are the i^{th} input and the r^{th} output of DMU_k ($k = 1, \dots, n$), respectively, λ_k is the weight allocated to DMU_k in constructing its ideal benchmark, $s_{i,k}^- \in \mathbb{R}^{m+}$ and $s_{r,k}^+ \in \mathbb{R}^{s+}$ denote the slacks of the first and second constraints; that is, input excesses and output shortfalls, and ρ_k is the SBM efficiency score of DMU_k . Let ρ_k^{CRS} , ρ_k^{IRS} , ρ_k^{DRS} and ρ_k^{VRS} denote the efficiency scores of DMU_k computed with each of the above mentioned SBM models.

Step 2: Use the following decision rules to determine the nature of the RTS for each DMU_k ($k = 1, \dots, n$):

- DMU_k operates on a CRS iff $\rho_k^{CRS} = \max\{\rho_k^{CRS}, \rho_k^{IRS}, \rho_k^{DRS}, \rho_k^{VRS}\}$
- DMU_k operates on a IRS iff $\rho_k^{IRS} = \max\{\rho_k^{CRS}, \rho_k^{IRS}, \rho_k^{DRS}, \rho_k^{VRS}\}$
- DMU_k operates on a DRS iff $\rho_k^{DRS} = \max\{\rho_k^{CRS}, \rho_k^{IRS}, \rho_k^{DRS}, \rho_k^{VRS}\}$
- DMU_k operates on a VRS iff $\rho_k^{VRS} = \max\{\rho_k^{CRS}, \rho_k^{IRS}, \rho_k^{DRS}, \rho_k^{VRS}\}$

These decision rules were proposed by Wu and An (2013) following the method proposed by Kerstens and Eeckaut (1999). One of the attractive features of this SBM based approach to the identification of the RTS status is that the projection of an inefficient *DMU* onto the efficient frontier takes account of both excesses in inputs and shortfalls in outputs.

Note that depending on whether IRS, DRS or VRS condition prevails, one must add $\sum_{j \in J^\lambda} \lambda_j \geq 1$, $\sum_{j \in J^\lambda} \lambda_j \leq 1$ or $\sum_{j \in J^\lambda} \lambda_j = 1$, respectively, as an additional constraint to the linear programming (LP) models 1, 2 and 3 below.

Stage 2 – Classification of DMUs: Use the following algorithm to identify several levels of efficiency or several efficient frontiers (evaluation contexts), say L :

Step 1: Set the performance level counter, say ℓ , equal to 1. Let $J^\ell = \{DMU_k, k = 1, \dots, n\}$ be the set of all n DMUs at efficiency level ℓ . Evaluate the entire set of DMUs, J^ℓ , by solving the relevant DEA model to construct the ℓ -level efficient frontier, say E^ℓ , where $E^\ell = \{k \in J^\ell | DEA \text{ score } \rho_k^\ell = 1\}$.

Step 2: Drop the current efficient DMUs; that is, E^ℓ , from the next DEA analysis run; that is, set $J^{\ell+1} = J^\ell - E^\ell$, and increase the counter ℓ by 1.

Step 3: Evaluate the “globally inefficient” set of DMUs identified in the previous step; that is, J^ℓ , by solving the relevant DEA model and set the current ℓ -level efficiency frontier to E^ℓ .

Step 4: If $J^\ell = \emptyset$, then stop; otherwise, set $\ell = \ell + 1$ and go to step 2.

where the relevant DEA model to determine the ℓ^{th} –level efficiency frontier is the slacks-based measure (SBM) model of Tone (2001b):

$$\begin{aligned}
 \text{Min} \quad & \rho_k^\lambda = \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{s_{i,k}^-}{x_{i,k}}\right) / \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_{r,k}^+}{y_{r,k}}\right) \\
 \text{s.t.} \quad & \sum_{j \in J^\lambda} \lambda_j x_{i,j} + s_{i,k}^- = x_{i,k}; \forall_i \\
 & \sum_{j \in J^\lambda} \lambda_j y_{r,j} - s_{r,k}^+ = y_{r,k}; \forall_r \\
 & \lambda_i \geq 0; \forall_j \in J^\lambda; s_{i,k}^- \geq 0, \forall_i; s_{r,k}^+ \geq 0, \forall_r
 \end{aligned} \tag{1}$$

where ρ_k^ℓ is the SBM efficiency score of DMU_k with respect to evaluation context ℓ . In the case that the optimal value of $\rho_k^\ell = 1$, then DMU_k is part of ℓ -level efficient frontier; otherwise DMU_k is inefficient and will be evaluated in future DEA runs. Obviously, DMUs are partitioned into L efficient frontiers, which indicate different performance levels. One could rank order DMUs considering the 1st-level efficient frontier DMUs as best and the L^{th} -level efficient frontier DMUs as worst, however, ties exist between DMUs on the same level efficient frontier and the next stage is designed to break those ties.

Stage 3 – Breaking of Efficiency Ties: Perform the following steps to break the ties between DMUs in the same level efficient frontier:

Step 1: Solve the LP (2) for all DMUs obtained at performance level ℓ ; that is, all $DMU_k \in E^\ell$, where $\ell = 2, 3, \dots, L$, to compute *relative progress scores*, δ_k^1 , with reference to the best evaluation context, E^1 , and rank DMUs on efficient frontier E^ℓ based on the calculated scores:

$$\text{Min } \delta_k^1 = \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{t_{i,k}^-}{x_{i,k}}\right) / \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{t_{r,k}^+}{y_{r,k}}\right) \quad (2)$$

$$s. t. : \sum_{j \in E^1} \lambda_j x_{i,j} \geq x_{i,k} - t_{i,k}^-; \forall i$$

$$\sum_{j \in E^1} \lambda_j y_{r,j} \leq y_{r,k} + t_{r,k}^+; \forall r$$

$$\lambda_i \geq 0; \forall j \in E^1; t_{i,k}^- \geq 0, \forall i; t_{r,k}^+ \geq 0, \forall r$$

where $t_{i,k}^-$ (respectively, $t_{r,k}^+$) indicates the amount by which input i (respectively, output r) of DMU_k should be decreased (respectively, increased) to reach the evaluation context E^1 .

Step 2: Solve the LP (3) for all DMUs obtained at the best efficient frontier E^1 ; that is, $DMU_k \in E^1$, to compute *relative attractiveness scores*, γ_k^2 , with reference to the second best evaluation context, E^2 , and rank DMUs on the best efficient frontier E^1 , based on the calculated scores.

$$\text{Max } \gamma_k^2 = \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{t_{i,k}^+}{x_{i,k}}\right) / \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{t_{r,k}^-}{y_{r,k}}\right) \quad (3)$$

$$s. t. : \sum_{j \in E^2} \lambda_j x_{i,j} \leq x_{i,k} + t_{i,k}^+; \forall i$$

$$\sum_{j \in E^2} \lambda_j y_{r,j} \geq y_{r,k} - t_{r,k}^-; \forall r$$

$$\lambda_i \geq 0; \forall j \in E^2; t_{i,k}^- \geq 0, \forall i; t_{r,k}^+ \geq 0, \forall r$$

where $t_{i,k}^+$ (respectively, $t_{r,k}^-$) indicates the amount by which input i (respectively, output r) of $DMU_k \in E^1$ should be increased (respectively, decreased) to reach to evaluation context E^2 .

In section 5, we shall use the above-described methodology to rank order competing corporate distress prediction models and discuss the empirical results obtained using UK data on firms listed on the London Stock Exchange (LSE) for the period 2008-2014. In this paper, DMUs are thirty competing corporate distress prediction models – see, Appendix A for a general description of these models. The inputs and outputs are the performance measures of the relevant criteria for assessing corporate prediction models. This study considers discriminatory power, calibration accuracy, information content, and correctness of categorical predictions criteria and their measures. Further, inputs (respectively, outputs) are selected based on the rule of the less (respectively, the more), the better; therefore, inputs (respectively, outputs) refer to the performance measures to be minimised (respectively, maximised).

To conclude this section, we would like to stress out that, from an empirical perspective, the proposed analysis framework could be used without any concerns as long as the relationship “#DMUs $\geq 2(\text{\#inputs} + \text{\#outputs})$ ” holds for the first performance level and the chosen evaluation context. In sum, the minimum sample size should satisfy the above condition for both the first performance level and the chosen evaluation context.

The above proposed methodology for ranking prediction models overcomes the limitations of the super-efficiency framework proposed by Mousavi et al (2015). This CDEA based methodology could be used to rank order prediction models. However, one could argue that the above mentioned CDEA methodology assumes that DMUs are homogeneous and therefore is not appropriate for assessing the prediction models under evaluation in our comparative analysis, because they belong to two different classes; namely, static models and dynamic models. In the next section, we propose a ranking methodology, which is suitable for ranking heterogeneous DMUs – whether prediction models, individuals, or organizations.

4. A Cross-Benchmarking-Cross-Efficiency Framework for Assessing Corporate Distress Predictions

In this section, we propose a ranking methodology that takes account of the heterogeneous nature of DMUs (e.g., prediction models, individuals, organizations). To be more specific, the proposed methodology is a hybrid framework that makes use of cross-benchmarking and cross-efficiency – referred to hereafter as CBCE, and could be summarised as follows:

Stage 1: Cross-Benchmarking Analysis

Cross-benchmarking evaluates each DMU_k ($k = 1, \dots, n$) against each cluster C_ℓ separately ($\ell = 1, \dots, \#C$) where $\#C$ denotes the number of clusters; in sum, as many DEA analyses as the number of clusters are performed each time with the reference set being a different cluster. Therefore, for each DMU_k ($k = 1, \dots, n$) and cluster C_ℓ ($\ell = 1, \dots, \#C$), solve the following variant of the SBM model of Tone (2001b) under the relevant RTS scheme; that is, CRS (i.e., variant of SBM model without any additional constraint), IRS (i.e., by augmenting the variant of SBM model with the constraint $\sum_{j \in C_\ell} \lambda_j^\ell \geq 1$), DRS (i.e., by augmenting the variant of the SBM model with the constraint $\sum_{j \in C_\ell} \lambda_j^\ell \leq 1$), and VRS (i.e., by augmenting the variant of the SBM model with the constraint $\sum_{j \in C_\ell} \lambda_j^\ell = 1$):

$$\begin{aligned}
 \text{Min} \quad & \rho_{k,k}^\ell = \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{s_{i,k}^{\ell-}}{x_{i,k}}\right) / \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_{r,k}^{\ell+}}{y_{r,k}}\right) \\
 \text{s.t.} \quad & \sum_{j \in C_\ell} \lambda_j^\ell x_{i,j} + s_{i,k}^{\ell-} = x_{i,k}; \forall i \\
 & \sum_{j \in C_\ell} \lambda_j^\ell y_{r,j} - s_{r,k}^{\ell+} = y_{r,k}; \forall r \\
 & \lambda_j^\ell \geq 0; \forall j \in C_\ell; s_{i,k}^{\ell-} \geq 0, \forall i; s_{r,k}^{\ell+} \geq 0, \forall r
 \end{aligned}$$

where the $x_{i,j}$ ($i = 1, \dots, m$) and $y_{r,j}$ ($r = 1, \dots, s$) are the i^{th} input and the r^{th} output of DMU_k ($k = 1, \dots, n$), respectively, λ_k^ℓ is the weight allocated to DMU_k in constructing its ideal benchmark when the reference set is C_ℓ , $s_{i,k}^{\ell-} \in \mathbb{R}^{m+}$ and $s_{r,k}^{\ell+} \in \mathbb{R}^{s+}$ denote the slacks of the first and second constraints; that is, input excesses and output shortfalls, and $\rho_{k,k}^\ell$ is the SBM efficiency score of DMU_k when benchmarked against C_ℓ - we shall refer to these scores as self-evaluation scores. Let $v_{i,k}^\ell$ and $u_{r,k}^\ell$ denote the dual variables corresponding to the first and

second sets of constraints; that is, $v_{i,k}^\ell$ (respectively, $u_{r,k}^\ell$) is the weight of input i (respectively, output r) assigned by DMU_k when benchmarked against DMUs in C_ℓ .

Stage 2: Cross-Efficiency Analysis

For each DMU_k ($k = 1, \dots, n$), use the input and output weights, $v_{i,j}^\ell$ and $u_{r,j}^\ell$, determined in the previous stage to compute its peer-evaluation scores, say $\rho_{k,j}^\ell$, as follows:

$$\rho_{k,j}^\ell = \frac{\sum_{r=1}^s u_{r,j}^\ell y_{r,k}}{\sum_{i=1}^m v_{i,j}^\ell x_{i,k}}; j = 1, \dots, n; j \neq k; \ell = 1, \dots, \#C.$$

These scores are in fact cross-efficiency scores. Then, compute a statistic, say Δ_k , that measures the deviation of the self-appraised scores from the peer-appraised scores as follows:

$$\Delta_k = \frac{\bar{\rho}_{k,k} - \bar{\rho}_{k,j}}{\bar{\rho}_{k,j}}, \text{ where } \bar{\rho}_{k,k} = \frac{\sum_{\ell}^{\#C} \rho_{k,k}^\ell}{\#C} \text{ and } \bar{\rho}_{k,j} = \frac{\sum_{\ell}^{\#C} \sum_{j \in C_\ell, j \neq k} \rho_{k,j}^\ell}{\#C \times (\sum_{\ell}^{\#C} \#C_\ell - 1)}.$$

Finally, rank order DMUs or prediction models in ascending order of their Δ_k s.

In the next section, we shall use this methodology to rank order competing corporate distress prediction models and discuss the empirical results obtained using UK data on firms listed on the London Stock Exchange (LSE) for the period 2008-2014 under the same setup used for CDEA.

5. Empirical Investigation

This section summarizes our empirical investigation related decisions. In sum, hereafter, we provide the details on our dataset and sampling (see, section 5.1), features selection (see, section 5.2), model evaluation criteria and measures (see, section 5.3), and distress prediction models (see, section 5.4).

5.1 Data and sampling

We took the following steps to select our data set. First, we excluded financial and utility UK companies to avoid having a biased sample due to these categories firms being regulated and as such their financial statements would follow specific regulations. Then, we considered all the remaining companies listed on the London Stock Exchange (LSE) at any time during a 10-year period from 2005 through 2014. Second, we excluded the firms that are listed for less than two years on LSE, as historical information is a requirement for some modelling frameworks. Third, we excluded the firms with missing values for the main accounting information (e.g., sales, total assets) and market information (e.g., price), which are essential items for calculating many financial ratios (Lyandres and Zhdanov, 2013). We replaced the remaining missing values with

the recently observed ones for each firm (Zhou, 2013; Shumway, 2001). Fourth, for each variable, we winsorised the outliers by replacing the values higher (respectively, lower) than 99th (respectively, 1st) percentile with that 99th (respectively, 1st) percentile value (Shumway, 2001).

In this research we considered distress as failure event. With respect to the classification of firms into distressed and non-distressed, we followed the definition of Pindado et al. (2008) where a company is classified as distressed if it experiences both of the following conditions for two consecutive years: (1) its earnings before interest, taxes, depreciation and amortization (EBITDA) is lower than its interest expenses, and (2) it shows negative growth in market value. To be more specific, the distress variable, say y , equals 1 for financially distressed companies and equals 0 otherwise.

Since our pool of distress prediction models consists of both static and dynamic models, we implemented them and tested their performance out-of-sample within a rolling horizon framework to reduce the degree of heterogeneity between static and dynamic models. In addition, implementing the above-mentioned models within a rolling horizon framework would reduce any bias due changes in macroeconomic conditions. To this end, we used 3-year firm-year observations from year $t - n + 1$ to year t , ($n = 3$ and $2008 < t < 2013$), as training samples to fit models; that is, to estimate their coefficients. Then, we used the fitted models to predict distress in year $t + 1$. Therefore, we considered six 3-year training samples to fit the models and six 1-year hold-out samples to test the models. Table 2 presents the sample sizes.

[Insert *Table 2* Here]

5.2 Feature Selection

To select proper features for prediction models, we applied the following steps. First, we reviewed the literature to select the most commonly used features in other studies (e.g., Hebb, 2016; du Jardin, 2015; Zhou, 2014, 2013; Ravi Kumar and Ravi, 2007), where we end up with 83 accounting-based ratios and seven market-based items. Second, we used *t-tests* to choose features which show a significant difference between the means of two groups of distressed and healthy firms (Shin and Lee, 2002; Huang et al., 2004; Shin et al., 2005). Third, for further reduction of features, we applied factor analysis, and principal component analysis with VARIMAX technique (Chen, 2011, Mousavi et al., 2015). To be more specific, we used factor

analysis to select the variables that both the absolute values of their loadings and communalities are greater than 0.5 and 0.8, respectively. Finally, 34 variables, which presented high factor loadings and high communality values, were retained as input features into a stepwise procedure in each statistical framework.

[Insert *Table 3* Here]

5.3 Corporate Distress Models to be Assessed

In this paper, we organised models into three categories; namely, original models, refitted models and new models. In case of original models, we compare the performance of the most cited statistical, probability and stochastic models in the literature on distress prediction. To be more specific, we consider the MDA models proposed by Altman (1968), Altman (1983), Lis (1972), and Taffler (1984); the logit model proposed by Ohlson (1980); the probit model proposed by Zmijewski (1984); and the linear probability model proposed by Theodossiou (1991); the contingent claim analysis models proposed by Bharath and Shumway (2008), Hillegeist et al. (2004) and Jackson and Wood (2013); and the survival analysis model proposed by Shumway (2001). In the case of refitted models, we keep the explanatory variables of each original model and refit them with our new data set. On the other hand, in the case of new models, we develop new distress prediction models using different static and dynamic frameworks and fit them with our new data set. The static frameworks used in our study are MDA, logit, probit and linear probability analysis. The dynamic frameworks used in our study are duration-independent with (or without) time-independent baseline hazard rate, and different duration-dependent models, which contain a variety of time-varying baseline hazard rates.

Note that depending on the existence and specification of baseline hazard rate in dynamic or duration models, one could classify them into two subcategories; namely, duration-independent and duration-dependent frameworks (Nam et al., 2008). The duration independent models could be classified into duration-independent with time-independent baseline (DIWTIB) and duration independent without baseline (DIWOB). The difference between these two types of models is that the former one contains a constant baseline hazard rate, while the latter one does not contain baseline hazard rate (e.g., Shumway, 2001). On the other hand, the duration-dependent framework contains a time-dependent baseline hazard rate, as mentioned in Beck et al. (1998), who use time dummies to proxy the baseline hazard rate.

Since the use of time dummies as an indirect proxy for the baseline rate is less efficient, we follow Nam et al. (2008) and Gupta et al. (2015) in using time-varying features to proxy the time-dependent baseline rate. Therefore, taking into account the duration dependent (DD) framework, the models differ based on the type of baseline hazard rates, i.e., $\ln(\text{age})$, $1/\ln(\text{age})$, last year probability of distress (LPD), and volatility of exchange rate (VEX) – see, Appendix A for more details on models. In addition, considering Cox hazard model, I followed Kim and Partington (2014) in estimating the time-dependent baseline rate using the historical information of the firm. I refer to this model as duration dependent with firm's specific baseline rate (DDWFSB).

Considering 31 static and dynamic frameworks and 6 training samples, we ended up with 186 models including original, refitted and newly developed. The original models, refitted models, and new models are presented in Figure 2, Figure 3, and Figure 4 with white, grey, and black shapes, respectively, and static and dynamic models with a circle and non-circle shapes, respectively – see legends of Figure 2, Figure 3, and Figure 4. Table 3 presents the list of features used to develop the new models. Also, Table 4 presents the new models fed with 3-year training sample from 2011 to 2013. See Appendix A for more details on models.

In the next section, we shall assess the relative performance of these models implemented within a rolling horizon framework under both a single criterion and multiple criteria and their measures using the proposed DEA framework (see section 6.2).

[Insert *Table 4* Here]

5.4 Performance Criteria and Measures

Our objective of this study is to evaluate the relative performance of distress prediction models using UK data. For this, we follow Mousavi et al. (2015) to assess the performance of different models under four commonly used criteria; namely the discriminatory power, the calibration accuracy, the information content, and the correctness of categorical prediction. On the discriminatory power criterion, we use Receivable Operating Characteristic (ROC), Kolmogorov-Smirnov (KS) statistics, Gini Index (GI), and Information Value (IV) to measure how much a model is capable of discriminating between the distressed firms and the healthy ones. On the calibration accuracy criterion, we use Brier Score (BS) to measure how much a model is qualified in estimating the probability of distress (PD). On the information content

criterion, we follow Agarwal and Taffler (2008) and use a log-likelihood statistic (LL) and pseudo- R^2 to measure the extent to which the output of a model (e.g., PD, scores) carries enough information for prediction. Finally, with respect to the correctness of categorical prediction criterion, we use Type I errors (T1), Type II errors (T2), misclassification rate (MR), sensitivity (Sen), specificity (Spe), and overall correct classification (OCC) to measure how often a model can predict distressed firms (respectively, healthy firms) as distressed (respectively, healthy) ones. Figure 1 displays the whole process of designing and assessing a prediction model.

[Insert *Figure 1* Here]

6. Empirical Results

In this section, we organise our analyses into mono-criterion analysis (see section 6.1), multi-criteria analysis using CDEA (see section 6.2), multi-criteria analysis using CBCE (see section 6.3), and summarise our main findings.

6.1 Mono-Criterion Analysis

Figure 2 presents the mono-criterion (unidimensional) rankings of 31 models considering the average performance measures of models over a 7-year period from 2008 to 2014. For our data set, mono-criterion rankings results could be summarised as follows. First, regarding the performance of all competing models in our study, the new developed models outperform original models and refitted models. This finding suggests that the change in trend of information during time, as someone would expect, tend to affect the performance of corporate distress prediction models; therefore, out-dated original models or refitted original models with new data set do not seem to be as efficient as new models with respect to most of the performance measures.

[Insert *Figure 2* Here]

Second, relating to the comparison of new dynamic models and new static models in our study, for most of the performance measures, the new dynamic models outperform static ones. To be more specific, on most performance measures – see, for example, T2, Spe, AUC, Gini, KS, BS, LL and Pseudo- R^2 , new dynamic models 27, 28 and 25 (DD_1n(age), DD_VEX and DIWTIB_1n(age), respectively) are superior to other models. However, considering the performance measures T2, MR, OCC, Spe and Stability index, new static model 23 (New PA) is

the best performer. In general, the density of new dynamic models amongst the top-ranking performers is an indicator of their superiority.

Third, contingent claim analysis (CCA) models (models 16, 17 and 18) are not amongst the best performers. The only exception is model 17 (Hillegeist et al., 2004), which is ranked third under T2, MR, OCC, and Spe; however, model 18 (Bharath and Shumway, 2008) seems to outperform other CCA models for most performance measures.

Fourth, regarding the performance of the original MDA models refitted (i.e., models 9, 10, 11 and 12), considering different measures, the rankings of models show inconsistency. Also, amongst the refitted regression models (i.e., models 13, 14 and 15), for most performance measures, the logit model 14 (refitted Ohlson, 1990) outperforms others. In general, considering refitted models, the discrete-time dynamic model 16 (Shumway, 2001) and the logit model 14 (Ohlson, 1990) outperform other refitted models, for most performance measures.

Last, but not the least, regarding the out-of-sample performance of the original models, the discrete-time dynamic model 8 (Shumway, 2001) and static models 6 and 5 (Ohlson, 1990; Theodossiou, 1991, respectively) seem to outperform other original models.

Much like typical outcomes in the existing literature, the rankings under mono-criterion are facing two main issues. Firstly, the rankings of models are different not only for measures under different criteria – see, for example, T1 under correctness of categorical prediction criterion and ROC under discriminatory power criterion, but also for measures under the same criterion; see, for example, OCC and MR under correctness of categorical prediction criterion or KS and ROC under discriminatory power criterion – as it is the case in Theodossiou (1991), Bandyopadhyay (2006), and Tinoco and Wilson (2013). Secondly, the models' rankings tend to have ties corresponding to some measures – see, for example, measures of T1 and Sen. Consequently, practitioners cannot make an informed decision about the best distress prediction model. To overcome these issues, we propose a multi-criteria ranking framework, namely SBM-CDEA, which not only provides a single ranking using multiple criteria at the same time but also breaks the possible ties in the ranking of competing models.

6.2 Multi-Criteria Analysis using CDEA

Figure 4 presents the multi-criteria (multidimensional) rankings of the above mentioned 31 models using SBM-CDEA. Further, Table 5 provides the efficient frontiers obtained with SBM-CDEA. Also, following Mousavi et al. (2015), we provide the rankings of models using SBM-super efficiency DEA, see Figure 3, to compare the performance of two multi-criteria assessment frameworks.

[Insert *Figure 3* Here]

[Insert *Figure 4* Here]

In our empirical investigation, RTS analysis revealed that VRS conditions hold and therefore an additional constraint (i.e., $\sum_{j \in J} \lambda_j = 1$) need to be added to linear programming models 1, 2 and 3. In addition, for our data set, multi-criteria rankings under SBM-super efficiency DEA and SBM-CDEA show considerable consistency in the rankings of top five models, for most of combinations of measures; however, they do not provide a general consistency in the rankings of all models.

Furthermore, under SBM-CDEA, the results could be summarised as follows. First, on the performance of all competing models in our study, the new developed models outperform the original models and the original models refitted.

Second, on the performance of dynamic and static models in our study, for most of the combinations of measures, the dynamic models outperform static ones. To be more specific, regardless of the combinations of measures, the dynamic models 28, 30 (DD_LPD and DD_1/ln(age), respectively) followed by models 25 and 27 (DIWTIB_ln(age) and DD_VEX, respectively) are amongst the top five best performers. The exceptional performance of the dynamic models seems to suggest that taking account of the time-varying nature of predictors pays off. Also, considering T1 or OCC as measures, the new static model 23 (New PA) outperform other dynamic and static models. The superiority of model 23 is due to its exceptional performance on T2 error. Though, for most of combinations of measures, static models 23 and 22 (new PA and new LA, respectively) are superior to other static once.

Third, with respect to CCA models – which are systematically amongst the worst ranked models, model 17 (Hillegeist et al., 2004) outperforms model 18 (Bharath and Shumway, 2008) and model 19 (Jackson and Wood, 2013), for all combinations of measures.

Forth, amongst the original MDA models refitted (i.e., models 9, 10, 11 and 12), for most of the combinations of measures, model 10 and 9 (refitted Lis, 1972) and refitted Altman (1968), respectively) outperform others. Further, amongst the refitted regression models (i.e., models 13, 14 and 15), for most performance measures, model 14 (refitted Ohlson, 1990) outperforms others. Also, the logit model 14 seems to be the best performer amongst all refitted models. Finally, considering the performance of original models, the dynamic model 8 (Shumway, 2001) followed by the static models 2 and 5 (Lis, 1972 and Theodosiou, 1991, respectively) are amongst the best performers.

We use four alternative measures, each represents one criterion, and several combinations of measures of the performance criteria to find out about the robustness of the multi-criteria rankings to the choice of measures. The empirical findings reveal that the multi-criteria rankings differ from the mono-criterion ones. The multi-criteria rankings do not show ties in the rankings of DPMs. The results suggest that the choice of the SBM-CDEA framework is an effective way to overcome inconsistency in the ranking of corporate DPMs.

6.3 Multi-Criteria Analysis using CBCE

With respect to the implementation of CBCE, we divided prediction models into two clusters; namely, the cluster of static models (C_1) and the cluster of dynamic models (C_2). Figure 5 summarizes the multi-criteria (multidimensional) rankings of the 31 models under comparison using CBCE. In a nutshell, empirical findings reveal that the relative performance of some prediction models, as suggested by the CBCE heterogeneous methodology, is consistent with the one suggested by CDEA homogenous methodology, whereas the relative performance of other prediction models is not. This difference in the relative performance of distress prediction models is due to the use of fundamentally different methodologies; namely, homogenous and heterogenous ones. Although in most applications the outcome of the heterogeneous methodology would be considered by practitioners more appropriate or reliable, we argue that when DMUs are prediction models – as opposed to individuals or organizations – the homogenous methodology is more appropriate. In fact, for prediction models and regardless of

their design features, all that matters is their predictive ability. Therefore, our recommendation for both academics and practitioners is to use a homogenous methodology for ranking prediction models.

7. Conclusion

Prediction of corporate distress and bankruptcy is one of the most crucial inputs to decisions making processes related to financing and investing activities. During the recent decades, academics and practitioners have developed many distress prediction models, which raise the question of “which of these models perform better in predicting distress?” To answer this question, the unidimensional ranking of competing prediction models has been the dominant approach; however, it results in conflicting rankings of models once someone shifts from one performance criterion to another. Mousavi et al. (2015) proposed a multi-criteria evaluation framework; namely, an orientation free super-efficiency DEA-based framework, to evaluate the performance of different bankruptcy prediction models, which provides a single ranking based on multiple performance criteria; such a framework faces one main issue that was overcome in this paper. In sum, we proposed an orientation-free slack-based context dependent DEA framework to overcome the methodological issues of both super-efficiency DEA-based and unidimensional ranking. Furthermore, we performed an exhaustive comparative analysis of the most popular distress modelling and prediction frameworks resulting in 31 prediction models including our models organised into three categories; namely, original models, original models refitted, and new models. We used several performance measures under four commonly used performance criteria, which are often employed in the literature to compare the performance of prediction models. A UK dataset on firms listed on the London Stock Exchange has been used to illustrate the proposed framework. The main findings could be summarised as follows. First, although the rankings of distress prediction models under non-oriented SBM-super efficiency and non-oriented SBM-CDEA are very similar, the latter one, however, does not suffer from the changes of reference benchmark from one DMU or prediction model to another. Second, our numerical results reveal that amongst the dynamic models, which are always superior in performance, duration dependent models (i.e., DD_VEX and DD_1/ln(age)), that use volatility of exchange rate (VEX) and 1/ln(age) as time-varying baseline, respectively, followed by duration independent models without baseline (i.e., DIWOB) tend to be superior. Third, numerical results seem to suggest that amongst the static models, LPA and PA models

outperform others. Last, but not the least, our empirical results suggest that developing new models using the most recent accounting, market, and macroeconomic information enhances the performance of distress prediction models.

However, one potential argument against the conventional mono-criterion and the proposed multi-criteria CDEA evaluation frameworks is that they assume implicitly or explicitly that DMUs are homogeneous whereas the prediction models under comparison are heterogeneous; that is, prediction models under comparison belong to difference categories characterised by different design features. Therefore, we proposed a new multi-criteria analysis framework that takes account of the heterogeneous nature of DMUs, which we refer to as CBCE, to evaluate the performance of competing static and dynamic distress prediction models. As expected, the rankings of models by CBCE are not always consistent with those provided by CDEA. While in many applications heterogeneous benchmarking methodologies deliver “fairer” efficiency profiles or rankings than homogeneous ones, in this specific application (i.e., assessing the relative performance of distress prediction models), using a heterogeneous benchmarking methodology is not suitable, because what matters is the predictive ability of models regardless of their design features. In addition, fairness is not an issue for prediction models as compares to other DMUs (e.g. individuals, firms, banks, universities) – no model will argue its ranking based on the limitations of its design features! In sum, our recommendation is to use a homogenous benchmarking methodology for assessing the relative performance of prediction models.

In practice, analysts could make use of many aspects of this research. First, instead of using a mono-criterion methodology to guide their choice of the prediction model or models to implement in actual applications, they could use the proposed multi-criteria CDEA framework, which is a more appropriate methodological choice both conceptually and in real-life decision environments and applications. As to the application; that is, corporate distress prediction, the high performance of the new models suggests that the more the relevant information categories these prediction models are fed with, the better is their predictive performance.

References

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance* 32(8), 1541–1551.

- Altman, E.I., Iwanicz- Drozdowska, M., Laitinen, E.K. and Suvas, A., 2017. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management & Accounting*, 28(2), 131-171.
- Altman, E., 1983. *Corporate Financial Distress: A Complete Guide to predicting, Avoiding and Dealing with Bankruptcy*. John Wiley and Sons.
- Altman, E.I., 1973. Predicting Railroad Bankruptcies in America. *The Bell Journal of Economics and Management Science* 4 (1), 184.
- Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance* 23 (4), 589–609.
- Avkiran, N.K. and Cai, L., 2014. Identifying distress among banks prior to a major crisis using non-oriented super-SBM. *Annals of Operations Research*, 217(1), 31-53.
- Bahrammirzaee, A., 2010. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19(8), 1165-1195.
- Balcaen, S. and Ooghe, H., 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), 63-93.
- Bandyopadhyay, A., 2006. Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches. *Journal of Risk Finance*, 7(3), 255–272.
- Banker, R.D., 1984. Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17(1), 35-44.
- Banker, R.D., Charnes, A. and Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), 1078-1092.
- Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M., Zhu, J., 2004. Returns to scale in different DEA models. *European Journal of Operational Research* 154 (2), 345–362.
- Barboza, F., Kimura, H. and Altman, E., 2017. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, pp.405-417.
- Basel Committee on Banking Supervision, 2000. *Principles for the management of credit risk*. Basel.
- Bauer, J., Agarwal, V., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance* 40(Mar), 432–442.
- Beaver, W.H., 1968. Alternative accounting measures as predictors of failure. *Accounting review* 43(1), 113–122.
- Beaver, W.H., 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research* 4, 71–111.
- Beck, N., Katz, J.N., Tucker, R., 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42, 1260–1288.

- Bharath, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. *The Review of Financial Studies* 21(3), 1339–1369.
- Bhimani, A., Gulamhussen, M.A. and Lopes, S.D.R., 2013. The role of financial, macroeconomic, and non-financial information in bank loan default timing prediction. *European Accounting Review*, 22(4), 739-763.
- Branch, B., 2002. The costs of bankruptcy: A review. *International Review of Financial Analysis*, 11(1), 39-57.
- Callejón et al., 2013, A.M. Callejón, A.M. Casado, M.A. Fernández, J.I. Peláez A system of insolvency prediction for industrial companies using a financial alternative model with neural networks, *International Journal of Computational Intelligence Systems*, 6 (1), 29-3.
- Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. *The Journal of Finance* 63(6), 2899–2939.
- Charalambous, C., Charitou, A. and Kaourou, F., 2000. Comparative analysis of artificial neural network models: Application in bankruptcy prediction. *Annals of operations research*, 99(1-4), 403-425.
- Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European journal of operational research* 2(6), 429–444.
- Chava, S., Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. *Review of Finance* 8(4), 537–569.
- Chen, N., Ribeiro, B. and Chen, A., 2016. Financial credit risk assessment: a recent review. *Artificial Intelligence Review*, 45(1), 1-23.
- Cooper, W.W., Seiford, L.M., Tone, K., 2006. *Introduction to data envelopment analysis and its uses*. Springer.
- Crapp, H.R., Stevenson, M., 1987. Development of a method to assess the relevant variables and the probability of financial distress. *Australian journal of management* 12(2), 221–236.
- Davydenko, S.A., Strebulaev, I.A. and Zhao, X., 2012. A market-based study of the cost of default. *The Review of Financial Studies*, 25(10), 2959-2999.
- Elkamhi, R., Ericsson, J. and Parsons, C.A., 2012. The cost and timing of financial distress. *Journal of Financial Economics*, 105(1), 62-81
- Fare, R., Grosskopf, S. and Lovell, C.K., 1994. *Production frontiers*. Cambridge University Press.
- Fethi, M.D. and Pasiouras, F., 2010. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European journal of operational research*, 204(2), pp.189-198.
- Geng, R., Bose, I. and Chen, X., 2015. Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, 241(1), pp.236-247.
- Gruber, M.J. and Warner, J.B., 1977. Bankruptcy costs: Some evidence. *The journal of Finance*, 32(2), 337-347.

- Gupta, J., Gregoriou, A., Healy, J., 2015. Using hazard models correctly: a comparison employing different definitions of SMEs financial distress.
- Hillegeist, S.A., Keating, E.K., Cram, D.P., Lundstedt, K.G., 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9(1), 5–34.
- Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems, Data mining for financial decision making* 37(4), 543–558.
- Jackson, R.H., Wood, A., 2013. The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review* 45(3), 183–202.
- Jiang, C., Wang, Z., Wang, R. and Ding, Y., 2017. Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Annals of Operations Research*, pp.1-19.
- Kerstens, K. and Eeckaut, P.V., 1999. Estimating returns to scale using non-parametric deterministic technologies: a new method based on goodness-of-fit. *European Journal of Operational Research*, 113(1), 206-214.
- Kim, H.J., Jo, N.O. and Shin, K.S., 2016. Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications*, 59, 226-234.
- Kou, G., Peng, Y. and Lu, C., 2014. MCDM approach to evaluating bank loan default models. *Technological and Economic Development of Economy*, 20(2), 292-311.
- Lane, W.R., Looney, S.W., Wansley, J.W., 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance* 10(4), 511–531.
- Li, Z., Crook, J. and Andreeva, G., 2014. Chinese companies distress prediction: an application of data envelopment analysis. *Journal of the Operational Research Society*, 65(3), pp.466-479.
- Li, Z., Crook, J. and Andreeva, G., 2017. Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications*, 80, 94-106.
- Liang, D., Lu, C.C., Tsai, C.F. and Shih, G.A., 2016. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561-572.
- Luoma, M., Laitinen, E.K., 1991. Survival analysis as a tool for company failure prediction. *Omega* 19(6), 673–678.
- Maddala, G.S., 1986. *Limited-dependent and qualitative variables in econometrics*. Cambridge university press.
- Martin, D., 1977. Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance* 1(3), 249–276.
- Meyer, P.A., Pifer, H.W., 1970. Prediction of bank failures. *The Journal of Finance* 25(4), 853–868.
- Morita, H., Hirokawa, K., Zhu, J., 2005. A slack-based measure of efficiency in context-dependent data envelopment analysis. *Omega* 33(4), 357–362.

- Mousavi, M.M., Ouenniche, J., Xu, B., 2015. Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis* 42, 64–75.
- Nam, C.W., Kim, T.S., Park, N.J., Lee, H.K., 2008. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *J. Forecast.* 27, 493–506.
- Neves, J.C., Vieira, A., 2006. Improving bankruptcy prediction with Hidden Layer Learning Vector Quantization. *European Accounting Review* 15, 253–271.
- Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research* 18(1), 109–131.
- Ouenniche, J., Xu, B., Tone, K., 2014. Relative performance evaluation of competing crude oil prices' volatility forecasting models: a slacks-based super-efficiency DEA model. *American Journal of Operations Research* 4(4), 235–245.
- Ouenniche, J. and Tone, K., 2017. An out-of-sample evaluation framework for DEA with application in bankruptcy prediction. *Annals of Operations Research*, 254(1-2), 235-250.
- Piesse, J., Lee, C.F., Kuo, H.C. and Lin, L., 2006. Corporate failure: Definitions, methods, and failure prediction models. In *Encyclopedia of Finance* (pp. 477-490). Springer US.
- Premachandra, I.M., Bhabra, G.S. and Sueyoshi, T., 2009. DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research*, 193(2), 412-424.
- Premachandra, I.M., Chen, Y. and Watson, J., 2011. DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega*, 39(6), 620-626.
- Reisz, A.S., Perlich, C., 2007. A market-based framework for bankruptcy prediction. *Journal of Financial Stability* 3(2), 85–131.
- Seiford, L.M., Zhu, J., 2003. Context-dependent data envelopment analysis—measuring attractiveness and progress. *Omega* 31(5), 397–408.
- Shin, K.-S., Lee, T.S., Kim, H., 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28(1), 127–135.
- Shin, K.-S., Lee, Y.-J., 2002. A genetic algorithm application in bankruptcy prediction modelling. *Expert Systems with Applications* 23(3), 321–328.
- Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74(1), 101-124.
- Sun, J., Fujita, H., Chen, P. and Li, H., 2017. Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. *Knowledge-Based Systems*, 120, 4-14
- Taffler, R.J., 1984. Empirical models for the monitoring of UK corporations. *Journal of Banking & Finance* 8(2), 199–227.
- Theodossiou, P., 1991. Alternative models for assessing the financial condition of business in Greece. *Journal of Business Finance & Accounting* 18(5), 697–720.

- Tinoco, M.H., Wilson, N., 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis* 30, 394–419.
- Tone, K., 2001a. On returns to scale under weight restrictions in data envelopment analysis. *Journal of Productivity Analysis*, 16(1), 31-47.
- Tone, K., 2001b. A slacks-based measure of efficiency in data envelopment analysis. *European journal of operational research*, 130(3), 498-509.
- Trujillo-Ponce, A., Samaniego-Medina, R., Cardone-Riportella, C., 2014. Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management* 15, 253–276.
- Wilson, R.L. and Sharda, R., 1994. Bankruptcy prediction using neural networks. *Decision support systems*, 11(5), 545-557.
- Wu, J. and An, Q., 2013. Slacks-based measurement models for estimating returns to scale. *International Journal of Information and Decision Sciences*, 5(1), 25-35.
- Wu, Y., Gaunt, C., Gray, S., 2010. A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics* 6(1), 34–45.
- Yeh, C.C., Chi, D.J. and Hsu, M.F., 2010. A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*, 37(2), pp.1535-1541.
- Zhou, L., 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems* 41, 16–25.
- Zhou, L., Lu, D. and Fujita, H., 2015. The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches. *Knowledge-Based Systems*, 85, pp.52-61.
- Zhu, J., 2014. Context-dependent Data Envelopment Analysis, in: *Quantitative Models for Performance Evaluation and Benchmarking*, International Series in Operations Research & Management Science. Springer International Publishing, pp. 153–174.
- Zmijewski, M.E., 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22, 59–82.

Figure 1: The Process of Designing and Assessing Distress Prediction Models

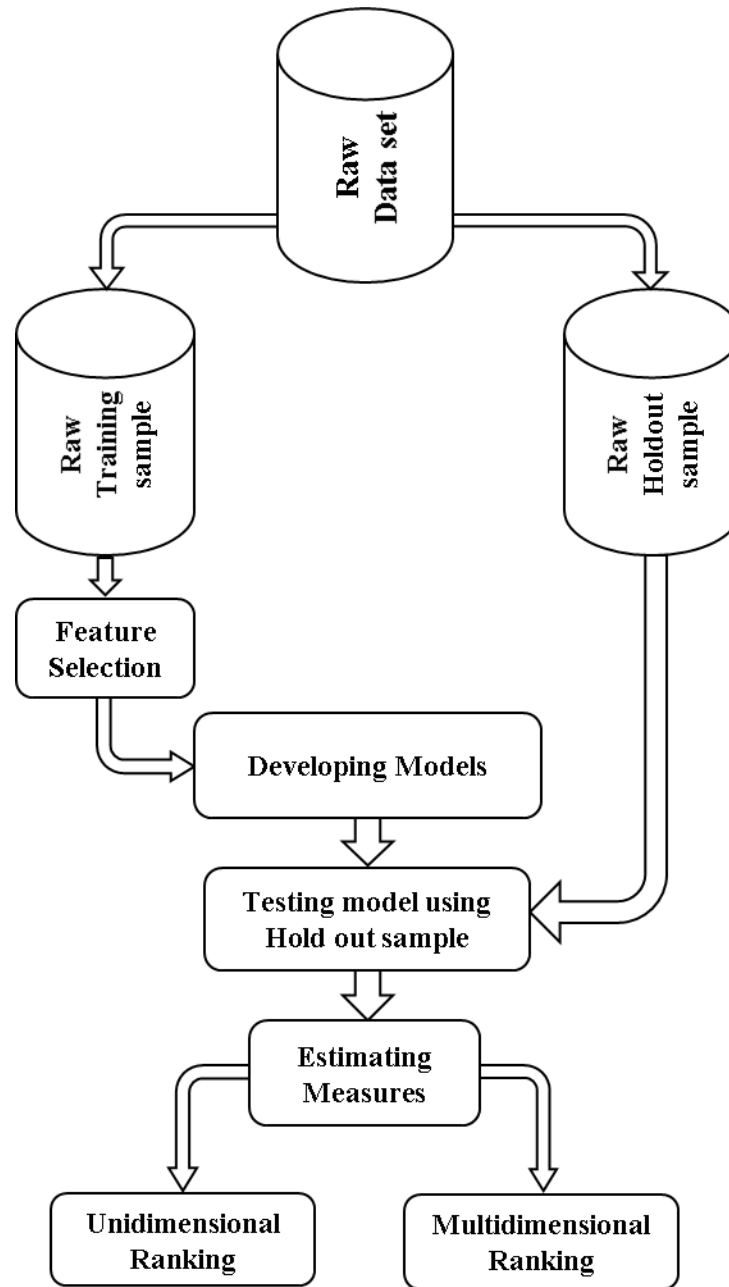


Figure 2: Mono-Criterion Rankings of Corporate Distress Prediction Models

This table presents the mono-criterion rankings of 31 competing corporate distress models, where models are ranked from best to worst using a single measure of a single criterion at a time. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; ROC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo- R^2 (R^2) are used as measures of information content. Circle shapes represent static models, namely Multivariate Discriminant Analysis (MDA), Linear Probability (LPA), Logit Analysis (LA), and Probit Analysis (PA). Non-circle shapes represent dynamic models, namely duration models, and Contingent Claim Analysis (CCA) models. White, grey, and black coloured shapes represent the original models, the original models refitted, and the new models, respectively.

Measure	Rank from the Best to Worst
T1; Sen	30 8 24 22 21 26 16 10 20 28 27 14 11 9 25 4 6 29 5 3 1 13 2 7 31 15 18 19 12 17 23
T2; MR; OCC; Spe	23 28 17 27 26 24 22 25 29 3 8 16 1 2 13 20 21 5 15 14 6 7 9 11 10 4 12 30 18 31 19
AUC; Gini	28 27 24 29 30 26 22 23 8 25 16 21 20 14 10 31 6 9 11 5 4 7 13 1 3 15 18 2 12 17 19
KS	28 27 30 29 24 22 23 26 25 8 16 21 20 14 10 31 9 11 6 5 4 1 3 13 2 15 7 18 12 17 19
IV	25 26 30 23 24 29 28 27 31 16 21 22 6 14 5 3 1 20 2 4 7 18 9 11 8 10 13 19 12 15 17
BS	25 28 27 31 30 29 24 26 14 7 5 2 18 16 20 21 23 1 11 9 22 10 13 8 4 15 19 17 12 3 6
LL, Pseudo- R^2	28 27 29 25 30 24 22 23 31 26 16 14 21 20 8 18 13 9 11 10 5 17 15 3 6 1 7 19 12 2 4
Stability index	23 31 30 29 21 20 16 26 19 22 1 24 25 18 27 28 3 14 7 6 5 17 4 8 9 2 11 10 13 15 12

^{1,9} Altman (1968); ^{2,10} Lis (1972); ^{3,11} Altman (1983); ^{4,12} Taffler (1984); ^{5,13} Theodossiou (1991); ^{6,14} Ohlson (1990); ^{7,15} Zmijewski (1984); ^{8,16} Shumway (2001);

¹⁷ Hillegeist et al. (2004); ¹⁸ Bharath and Shumway (2008); ¹⁹ Jackson and Wood (2013); ²⁰ New MDA; ²¹ New LPA; ²² New LA; ²³ New PA; ²⁴ DIWOB;

²⁵ DIWTIB_ln(age); ²⁶ DD_ln(age); ²⁷ DD_VEX; ²⁸ DD_LPD; ²⁹ DIWTIB_1/ln(age); ³⁰ DD_1/ln(age); ³¹ DDWFSB

Figure 3: SBM-Super Efficiency DEA-based Multi-Criteria Rankings of Corporate Distress Prediction Models

This table presents the multi-criteria rankings of 31 competing corporate distress models using a DEA ranking framework, where models are ranked from best to worst using SBM-super efficiency scores. A multi-criteria ranking is produced for each combination of a variety of metrics of the performance criteria under consideration, where inputs (resp. outputs) are chosen according to the principle of the less (resp. more) the better. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; ROC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo- R^2 (R^2) are used as measures of information content. Circle shapes represent static models, namely Multivariate Discriminant Analysis (MDA), Linear Probability (LPA), Logit Analysis (LA), and Probit Analysis (PA). Non-circle shapes represent dynamic models, namely duration models, and Contingent Claim Analysis (CCA) models. White, grey, and black coloured shapes represent the original models, the original models refitted, and the new models, respectively.

Inputs	Outputs	Rank from the Best to Worst
T1; BS; LL	ROC	30 28 25 27 24 22 26 29 16 21 14 20 31 23 8 10 9 11 5 4 13 1 7 18 15 2 12 19 17 6 3
T2; BS; LL	ROC	23 28 25 27 29 24 26 30 22 31 14 16 21 20 8 5 10 9 11 7 18 13 2 15 1 4 17 12 19 6 3
BS; LL	ROC; OCC	23 28 25 27 29 24 22 26 30 16 14 31 21 20 8 5 10 9 11 13 15 18 1 7 17 2 4 12 19 3 6
T1; BS; LL	KS	30 28 25 27 24 22 26 29 16 21 14 20 23 31 8 10 9 11 5 4 1 13 2 15 7 18 6 12 17 3 19
T2; BS; LL	KS	23 28 25 27 29 24 26 30 22 31 14 16 21 20 8 5 10 9 11 4 1 13 2 7 15 18 12 17 19 6 3
T1; BS	ROC; R^2	30 28 25 27 24 22 26 29 16 14 21 20 31 23 8 18 11 9 13 10 5 15 17 1 7 19 12 6 3 4 2
T2; BS	ROC; R^2	23 28 25 27 29 24 26 30 31 14 22 16 21 20 8 18 13 9 11 10 5 17 15 7 1 19 12 2 3 4 6
BS	ROC; OCC; R^2	23 28 25 27 29 24 26 22 30 16 31 14 21 20 8 18 13 9 11 10 5 17 15 1 7 19 12 2 4 3 6
T1; BS	KS; R^2	30 23 28 25 24 22 27 26 8 16 21 20 29 14 10 9 11 5 4 1 13 2 7 31 15 18 17 12 19 6 3
T2; BS	KS; R^2	23 25 28 27 24 29 26 30 14 31 5 16 22 7 2 20 8 21 9 10 11 13 1 18 17 15 4 12 19 3 6

^{1,9} Altman (1968); ^{2,10} Lis (1972); ^{3,11} Altman (1983); ^{4,12} Taffler (1984); ^{5,13} Theodossiou (1991); ^{6,14} Ohlson (1990); ^{7,15} Zmijewski (1984); ^{8,16} Shumway (2001);

¹⁷ Hillegeist et al. (2004); ¹⁸ Bharath and Shumway (2008); ¹⁹ Jackson and Wood (2013); ²⁰ New MDA; ²¹ New LPA; ²² New LA; ²³ New PA; ²⁴ DIWOB;

²⁵ DIWTIB_ln(age); ²⁶ DD_ln(age); ²⁷ DD_VEX; ²⁸ DD_LPD; ²⁹ DIWTIB_1/ln(age); ³⁰ DD_1/ln(age); ³¹ DDWFSB

Figure 4: SBM-Context Dependent DEA-based Multi-Criteria Rankings of Corporate Distress Prediction Models

This table presents the multi-criteria rankings of 31 competing corporate distress models using a DEA ranking framework, where models are ranked from best to worst using SBM-CDEA scores. A multi-criteria ranking is produced for each combination of a variety of metrics of the performance criteria under consideration, where inputs (resp. outputs) are chosen according to the principle of the less (resp. more) the better. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; ROC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo-R² (R²) are used as measures of information content. Circle and non-circle shapes indicate static and dynamic frameworks, respectively. Black, grey and white shapes represent new models, BSM-based models, and original models refitted, respectively.

Inputs	Outputs	Rank from the Best to Worst
T1; BS; LL	ROC	30 28 25 27 24 8 22 26 29 31 16 21 14 23 20 10 5 7 9 11 18 2 6 4 13 1 15 17 3 12 19
T2; BS; LL	ROC	23 28 25 27 17 29 24 26 30 31 22 14 16 2 21 20 8 5 7 1 3 10 9 18 13 6 11 15 4 12 19
BS; LL	ROC; OCC	23 28 25 27 29 24 26 30 31 17 22 16 14 21 20 8 5 7 2 10 9 13 18 1 3 6 11 15 4 12 19
T1; BS; LL	KS	30 28 25 27 24 8 22 26 29 31 16 21 14 23 20 10 5 7 9 11 2 18 4 1 13 6 15 17 3 12 19
T2; BS; LL	KS	23 28 25 27 17 29 24 26 30 31 22 14 2 16 8 5 7 3 21 20 1 18 10 9 11 13 6 4 15 12 19
T1; BS	ROC; R ²	30 28 25 27 24 8 22 26 29 31 16 14 21 23 20 10 5 7 18 11 9 13 6 2 15 17 1 3 4 19 12
T2; BS	ROC; R ²	23 28 25 27 17 29 24 26 30 31 14 22 16 2 21 20 8 5 7 1 3 18 13 9 10 6 11 15 19 12 4
BS	ROC; OCC; R ²	23 28 25 27 29 24 30 26 31 17 22 16 14 21 20 8 18 5 7 2 13 9 11 10 1 3 6 15 4 19 12
T1; BS	KS; R ²	30 28 25 27 24 8 22 26 29 31 16 14 21 23 20 10 5 7 11 9 13 18 2 15 1 17 6 3 4 19 12
T2; BS	KS; R ²	23 28 25 27 17 29 24 26 30 31 22 14 2 16 8 5 7 3 20 21 18 1 13 9 11 10 6 15 4 19 12

^{1,9} Altman (1968); ^{2,10} Lis (1972); ^{3,11} Altman (1983); ^{4,12} Taffler (1984); ^{5,13} Theodossiou (1991); ^{6,14} Ohlson (1990); ^{7,15} Zmijewski (1984); ^{8,16} Shumway (2001);

¹⁷ Hillegeist et al. (2004); ¹⁸ Bharath and Shumway (2008); ¹⁹ Jackson and Wood (2013); ²⁰ New MDA; ²¹ New LPA; ²² New LA; ²³ New PA; ²⁴ DIWOB;

²⁵ DIWTIB_In(age); ²⁶ DD_In(age); ²⁷ DD_VEX; ²⁸ DD_LPD; ²⁹ DIWTIB_1/ln(age); ³⁰ DD_1/ln(age); ³¹ DDWFSB

Figure 5: CBCE Multi-Criteria Rankings of Corporate Distress Prediction Models

This table presents the multi-criteria rankings of 31 competing corporate distress models using the proposed CBCE ranking framework, where models are ranked from best to worst using CDEA scores. A multi-criteria ranking is produced for each combination of a variety of metrics of the performance criteria under consideration, where inputs (resp. outputs) are chosen according to the principle of the less (resp. more) the better. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; ROC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo- R^2 (R^2) are used as measures of information content. Circle and non-circle shapes indicate static and dynamic frameworks, respectively. Black, grey and white shapes represent new models, BSM-based models, and original models refitted, respectively.

T1; BS; LL	ROC	25 31 16 8 21 20 14 18 9 11 5 1 2 7 27 28 29 17 22 19 13 23 10 15 26 30 4 12 24 3 6
T2; BS; LL	ROC	25 31 16 23 18 20 21 22 14 5 7 2 8 19 17 13 28 27 9 11 24 15 10 29 1 30 4 12 6 3 26
BS; LL	ROC; OCC	25 21 20 14 22 31 16 5 7 2 18 29 30 24 27 28 23 13 10 9 11 15 6 12 1 4 3 26 8 17 19
T1; BS; LL	KS	25 31 16 21 20 14 8 9 11 5 1 26 2 7 18 23 27 28 29 22 13 10 15 30 24 4 12 17 19 3 6
T2; BS; LL	KS	25 23 31 16 26 18 21 20 22 14 5 2 7 17 8 13 9 10 11 15 19 1 4 12 30 24 28 27 29 3 6
T1; BS	ROC; R^2	2 23 4 18 17 19 7 5 16 31 25 12 14 21 20 8 1 15 6 3 10 9 11 13 30 28 27 24 29 26 22
T2; BS	ROC; R^2	2 4 18 19 17 7 16 5 23 31 25 8 21 14 20 12 1 6 15 10 9 11 3 13 28 27 29 24 26 30 22
BS	ROC; OCC; R^2	2 4 7 25 5 18 14 21 20 31 16 19 12 6 23 3 26 1 17 15 10 29 30 28 27 24 9 11 13 22 8
T1; BS	KS; R^2	31 2 4 8 18 16 1 5 7 11 9 26 14 21 20 19 25 17 23 10 12 24 30 15 22 28 27 13 29 3 6
T2; BS	KS; R^2	2 4 18 7 19 5 16 23 31 17 26 14 21 20 25 8 1 12 6 15 10 9 11 3 13 28 27 24 29 30 22

^{1,9} Altman (1968); ^{2,10} Lis (1972); ^{3,11} Altman (1983); ^{4,12} Taffler (1984); ^{5,13} Theodossiou (1991); ^{6,14} Ohlson (1990); ^{7,15} Zmijewski (1984); ^{8,16} Shumway (2001);

¹⁷ Hillegeist et al. (2004); ¹⁸ Bharath and Shumway (2008); ¹⁹ Jackson and Wood (2013); ²⁰ New MDA; ²¹ New LPA; ²² New LA; ²³ New PA; ²⁴ DIWOB;

²⁵ DIWTIB_ln(age); ²⁶ DD_ln(age); ²⁷ DD_VEX; ²⁸ DD_LPD; ²⁹ DIWTIB_1/ln(age); ³⁰ DD_1/ln(age); ³¹ DDWFSB

Table 1: Literature on Comparative Studies of Failure Prediction Models

The table present the studies that compared the performance of competing statistical failure (i.e. distress or bankruptcy) prediction models. All comparative studies used a mono-criterion framework, i.e. one measure of a criterion is applied at a time, to evaluate the performance of competing models. Then, the **conflict** in final rankings are seen in most of studies.

Author	Models	Criteria (Measure)	Result
<i>Panel I: Comparison between traditional statistical models</i>			
Press and Wilson (1976)	LA and MDA models	Correctness of categorical prediction (T1 and T2 errors)	Two models unlikely will give significantly different results.
Collins and Green (1982)	LPA, MDA, and LA models	Correctness of categorical prediction (OCC, T1 and T2)	The models produce identical, uniformly results.
Lo (1986)	MDA and LA models	Power of models	There are no differences between models.
Theodossiou (1991)	LPA, LA, and PA models	Correctness of categorical prediction (T1 and T2 errors), Calibration (BS), Information content (pseudo- R^2)	logit model outperforms others; conflict in the ranking of others on different measures
Lennox (1999)	LA, PA, and MDA models	Correctness of categorical prediction (T1 and T2)	A well-specified non-linear PA and LA are superior over DA
Bandyopadhyay (2006)	MDA models and logit models	Correctness of categorical prediction (OCC, T1 and T2) Discriminatory power (ROC), Information content (pseudo- R^2 , LL)	Conflict in rankings using different criteria and measures
Tinoco and Wilson (2013)	logit models taking to account different categories of features	Discriminatory power (ROC, Gini, KS), Calibration accuracy (HL)	Conflict in rankings using different criteria and measures
<i>Panel II: Comparison between traditional statistical models and survival analysis models</i>			
Luoma and Laitinen (1991)	Cox-hazard, MDA and LA models	Correctness of categorical prediction (T1 and T2)	SA model is inferior to MDA and LA models
Shumway (2001)	Discrete-time SA model, MDA, LA and PA	Correctness of categorical prediction (OCC)	Conflict in results. SA model which encompasses both accounting and market information (respectively, only accounting information) outperforms (respectively, underperforms) other statistical techniques.

Panel III: Comparison between statistical models and contingent claim analysis (CCA) models

Hilligeist et al. (2004)	BSM-based, LA and MDA models	Information content (LL and Pseudo-R ²)	BSM-based model outperforms both original and refitting version of LA and MDA models
Reisz and Perlich (2007)	BSM-based, KMV, DOC and MDA models	Discriminatory power (AUROC)	DOC and MDA outperforms others for 3-, 5- and 10-year ahead; MDA outperforms others for 1-year ahead distress prediction
Agarwal and Taffler (2008)	Contingent claim based models [HKCL (2004) and BHSH (2008)] and MDA model of Taffler (1984)	Discriminatory power (ROC), Information content (pseudo-R ² , LL), Correctness of categorical prediction (EV for different cost of misclassification)	MDA model outperforms HKCL (2004) on ROC and pseudo-R ² . Conversely , HKCL (2004) outperforms BHSH (2008) and MDA model on LL.
<i>Panel IV: Comparison between CCA models and survival analysis models</i>			
Campbell et al. (2008)	A new duration dependent SA without time-variant baseline, SA model [Shumway (2001)] and KMV (Kealhofer, McQuown and Vasicek) model	Information content (pseudo-R ² , LL)	The suggested new SA model outperforms both Shumway (2001) and KMV models.
<i>Panel V: Comparison between CCA, survival analysis and traditional statistical models</i>			
Wu et al. (2010)	MDA [Altman (1968)], Logit model [Ohlson (1980)], probit model [Zmijewski (1984)] hazard model [Shumway (2001)] and BSM- model [HKCL (2004)]	Information content (pseudo-R ² , LL) Correctness of categorical prediction (OCC), Discriminatory power (ROC)	Conflict in rankings. Shumway (2001) outperforms others on LL and Pseudo-R ² . Logit model performs better than others on OCC.
Bauer and Agarwal (2014)	Traditional model, contingent claim based model and hazard model	Discriminatory power (ROC), Information content (LL, R ²) and Correctness of categorical prediction (OCC, T1, T2)	Hazard model outperforms others; Conflict in ranking of others on different measures

Table 2: The Proportion of Distress Firms (*D*) in Training and Holdout Samples

This table presents the yearly proportion of distress in training and hold-out samples. The percentage of distress is presented based on the definition of distress (*D*).

Hold-out Sample					Training Sample				
Year	# Healthy	# Distress	Total	D%	Period	# Healthy	# Distress	Total	D%
2008	1,704	106	1,810	5.86%	2005-2007	5,423	175	5,598	3.13%
2009	1,456	165	1,621	10.18%	2006-2008	5,367	245	5,612	4.37%
2010	1,409	61	1,470	4.15%	2007-2009	4,986	352	5,338	6.59%
2011	1,354	27	1,381	1.96%	2008-2010	4,569	332	4,901	6.77%
2012	1,255	69	1,324	5.21%	2009-2011	4,219	253	4,472	5.66%
2013	1,143	101	1,244	8.12%	2010-2012	4,018	157	4,175	3.76%
2014	1,120	66	1,186	5.56%	2011-2013	3,752	197	3,949	4.99%
Total				5.66%*					5.03%*

* The number is the average.

Table 3: List of financial ratios

Category	Ratio or item	Category	Ratio or item
Profitability (9)	Net income to total liabilities	Liquidity (9)	Current asset turnover
	EBIT to total assets		Current assets to total liabilities
	Return on assets		Current liabilities to current assets
	Operating income after depreciation to total assets		Inventory to current assets
	Retained earnings to total assets		Inventory turnover
	Expected return on assets		Inventory to total assets
	Total liabilities exceed total assets		Profit before tax to current liabilities
	Changes in net income in two consecutive years		Quick asset to total assets
	Negative net income for last two years		Quick asset to inventory
Asset utilisation (2)	Asset turnover ratio	Solvency (3)	Current liabilities to liabilities
	Quick assets to sales		Equity to capital
Cash flow (2)	Operating cash flow to liabilities		Long term and current liabilities to total assets
	Funds provided by operations to total liabilities	Market information (5)	Lag of excess return
			Lag sigma
Mixed (2)	GDP \times Sales		Ln (price)
	Interest rate \times Income		Real size
			Failure rate in last year
		Firm characteristics (2)	Ln(age)
			Log (total assets to GNP price level index)

Table 4: New Designed Models

The table presents the features and coefficients of the new models (models 20 to 31 in our rankings), namely MDA (20), LPA (21), LA (22), PA (23), DIWOB (24), DIWTIB_{ln(age)} (25), DIWTIB_{1/ln(age)} (26), DDWTD_{ln(age)} (27), DDWTD_{VEX} (28), DDWTD_{LPD} (29), DDWTD_{1/ln(age)} (30) and DDWFSB (31). *** and ** refer to 1% and 5% significance level, respectively.

Models	Model 20	Model 21	Model 22	Model23	Model 24	Model 25*	Model 26*	Model 27	Model 28	Model 29	Model 30	Model 31
Explanatory variables												
Intercept	-2.94	0.073	-3.44	-3.24	-1.53	-1.53 $+ln(age)_i$	-1.53 $+1/ln(age)_i$	-1.9749	-0.2996	-2.171	-0.591	
Current Assets to Total Liabilities	-0.002	-0.0001*										
Net income to long term funding	-0.014	-0.0006***										0.0022
Current Assets to Sales	-.0002	-0.0001***										
Total liabilities to Total Assets	0.066	0.0028***										-0.0079
Cash and equivalent to Sales	0.006	0.0002***										0.0003
Inventory to Assets												-0.9997
Equity to Sales	-0.0003	-0.0001***										
Lag of Excess Return	-1.289	-0.059***	-0.832***	-0.9322***	-0.987***	-0.987***	-0.987***	-1.001***	-1.041***	-1.033***	-1.016***	-0.985***
Lag of Sigma	2.865	0.117***										
Ln (price)	-3.842	-0.015***	-0.281***	-0.211***	-0.217***	-0.217***	-0.217***	-0.218***	-0.209***	-0.213***	-0.214***	-0.226***
Equity to Capital		0.003	0.045	0.0204								
Current Liabilities to Total Assets		0.0067	0.038	0.032								
Real size			-0.2562***	-0.1524**								-0.111
Ohlson size					-0.197***	-0.197***	-0.197***	-0.201***	-0.193***	-0.193***	-0.207***	
Interest rate \times Net Income	-0.0001	3.23	-0.0006***	-0.00006***	-0.00007***	-0.00007***	-0.00007***	-0.00007***	-0.00006***	-0.00006***	-0.00007***	-0.00004***
GDP \times Sales	-0.0001	-3.34	-0.0004***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
Ln (age)								0.185				
1/ ln (age)											-2.124***	
Volatility of Exchange Rate (VEX)									-0.711***			
Last year distress rate										15.319***		

* Note that in models 25 and 29, the $ln(age)$ and $1/ln(age)$ of firm i is added to intercept as the baseline hazard rate.

Table 5: Efficient Frontiers with Different Performance Levels

Efficient Frontiers	(D T1; BS, LL (O) ROC	(D T2; BS; LL (O) ROC	(D BS; LL (O) ROC; OCC	(D T2; BS; LL (O) KS	(D T1; BS; LL (O) KS	(D T1; BS (O) ROC; R ²	(D T2; BS (O) ROC; R ²	(D BS (O) ROC; OCC; R ²	(D T1; BS (O) KS; R ²	(D T2; BS (O) KS; R ²
E¹	{25,30,27,23}	{21,30,28,27,25,24,23}	{21,30,28,27,25,24,23}	{21,30,27,25,23}	{23,30,25,24,21}	{21,30,27,25,23}	{21,30,28,27,25,24,23}	{21,23,24,25,27,28,30}	{21,23,25,27,30}	{21,23,24,25,30}
E²	{26,28,29}	{26,22,29,17}	{26,22,29,17}	{26,28,29}	{28,26,27,29,17}	{26,28,29}	{26,22,29,17}	{26,22,29,17}	{26,28,29}	{17,26,27,28,29}
E³	{22,20,24,14,1,6}	{20,14}	{20,14}	{22,20,24,14,1,6}	{22,14}	{22,20,24,14,6,1}	{20,14}	{20,14}	{22,20,24,14,6,1}	{14,22}
E⁴	{2,16,10,9}	{16}	{16}	{2,10,9,16}	{20}	{16,9,10,2}	{18,8,6,9,11,10,15,2}	{16}	{16,9,10,2}	{20}
E⁵	{11,8,18}	{2,9,10,11,8,15,6,18}	{9,11,2,8,10,15,6,18}	{11,8,18,}	{16,10}	{18,8,11,17,19,3}	{19,3,1,13}	{18,8,6,9,11,10,15,2}	{18,8,11}	{10,16}
E⁶	{3,19,17}	{1,3,13,9}	{1,3,13,19}	{3,7,19}	{11,9,2,8,15,6,18}	{15,13,12,4}	{5,12,4}	{19,3,1,13}	{17,19,3}	{2,6,8,9,11,15,18}
E⁷	{13,15,12,4}	{5,12,4}	{5,12,4}	{15,13,12,4}	{1,3,13,19}	{5,7}	{7}	{5,12,4}	{15,13,12,4}	{1,3,13,19}
E⁸	{5,7}	{7}	{7}	{5,7}	{7}			{7}	{5,7}	{4,5,12}
E⁹										{7}

Appendix A: Statistical Models of Corporate Distress Prediction

Framework	Model	Explanation
Multiple discriminant analysis (MDA)	Altman (1968) $Z = 1.2 WCTA + 1.4 RETA + 3.3 EBITTA + 0.6 MVTL + 0.999 STA$ WCTA: Working capital / Total Assets; RETA: Retained Earnings / Total Assets; EBITTA: Earnings before interest and taxes / Total assets; METL: Market value of equity / Total Liabilities; STA: Sales / Total assets	<p>Assuming there are n groups, the generic form of DA model for group k is:</p> $z_k = f\left(\sum_{j=1}^p \beta_{kj} x_j\right)$ <p>where x_j is discriminant feature j, β_{kj} is the discriminant coefficient of feature j in group k, z_k represents the score of group k, and f is the linear or non-linear classifier that maps the scores, say $\beta^t x$, onto a set of real numbers. To compare DA models to other statistical models, we need to estimate the probability of failure, which is used as an input for estimating many measures of performance. For this, we follow Hillegeist et al. (2004) in using a logit link to calculate the probability of failure for companies:</p> $P(distress)_i = \frac{e^z}{1 + e^z}$
Multiple discriminant analysis (MDA)	Altman (1983) $Z = 0.717 WCTA + 0.847 RETA + 3.107 EBITTA + 0.42 BVTL + 0.998 STA$ WCTA: Working capital / Total Assets; RETA: Retained Earnings / Total Assets; EBITTA: Earnings before interest and taxes / Total assets; BVETL: Book value of equity / Total Liabilities; STA: Sales / Total assets	
Multiple discriminant analysis (MDA)	Lis (1972) $Z = 0.063 WCTA + 0.092 RETA + 0.057 EBITTA + 0.0014 NWTL$ WCTA: Working capital/ Total assets; EBITIA: Earnings before interest and taxes/ Total assets; METL: Market value of equity /Total liabilities; NWTA: Net wealth / Total assets	
Multiple discriminant analysis (MDA)	Taffler (1984) $Z = 3.2 + 2.5 CATL + 12.18 PBTCL + 0.029 NCI - 10.68 CLTA$ CLTA: Current liabilities/ Total assets; PBTCL: Profit before tax/ Current liabilities; NCI: Number of credit intervals as (quick assets - current liabilities) / ((sales - PBT - depreciation)/365); CATL: Current assets / Total liabilities	
Linear probability model (LPA)	Theodossiou (1991) $Z = -0.075 + 0.51 WCTA - 0.21 TDTA + 0.449 NITA + 0.663 RETA - 0.446 LTDTA$ WCTA: Working capital/Total assets; TDTA= Total debt/Total assets; NITA: Net income/Total assets; RETA= Retained earnings/ Total assets; LTDTA= Long term debt/Total assets	<p>The generic linear probability model (LPA) is a particular case of OLS regression and results in an estimate of probability of distress, the formula for which is as follows;</p> $P(distress)_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$
Logit analysis (LA)	Ohlson (1980) $\log \left[\frac{P_i}{1 - P_i} \right] = -1.32 - 1.43 WCTA + 6.03 TLTA - 2.37 NITA - 0.407 OSIZE - 1.83 FUTL + 0.0757 CLCA + 0.285 INTWO - 1.72 OENEG - 0.521CHIN$ WCTA: Working capital/Total assets; TLTA: Total liabilities/ Total assets; NITA: Net income/ Total assets; OSIZE = log (Total assets/GNP price-level index); FUTL: Funds from operations (operating income	<p>The generic model for binary variables could be stated as follows:</p> $\begin{cases} P(distress)_i = P(Y = 1) \\ P(distress)_i = G(\beta, X) \end{cases}$ <p>where Y denotes the binary response variable, X denotes the vector of features, β denotes the vector of coefficients of X in the model, and $G(\cdot)$ is a link function that maps a score $\beta^t x$ onto a probability. In practice, depending on the choice of link function, the type of probability model is determined. For example, the logit model</p>

	minus depreciation) / Total liabilities; <i>CLCA</i> : Current liabilities/ Current assets; <i>INTWO</i> = 1 if net income has been negative for the last 2 years, 0 otherwise; <i>OENEG</i> = 0 if total liabilities exceed total assets, 1 otherwise; $CHIN = (NI_t - NI_{t-1}) / (NI_t + NI_{t-1})$, where NI_t is the net income for the last period. The variable is thus a proxy for the relative change in net income.	(respectively, probit model) assumes that the link function is the cumulative logistic distribution, say θ (respectively, cumulative standard normal distribution, say N) function.
Probit analysis (PA)	Zmijewski (1984) $\log [Pt/(1 - Pt)] = 4.336 - 5.769 TLTA + 4.513 NITA - 0.004 CACL$ <i>NITA</i> : Net income/ Total assets; <i>TLTA</i> : total liabilities/ Total assets; <i>CACL</i> : Current assets/ Current liabilities	
Contingent claim analysis (CAA): Black-Scholes-Merton (BSM) Based Models	Hillegeist et al. (2004), Bharath and Shumway (2008) $P(distress_i) = N\left(-\frac{\ln\left(\frac{V_a}{L}\right) + (\mu - \delta - 0.5\sigma_a^2) \times T}{\sigma_a \sqrt{T}}\right)$ <i>N</i> (.): the cumulative normal distribution function, V_a : the value of the company's assets; <i>L</i> : total liabilities; μ : the expected return of the firm; σ_a : volatility of the company's asset; δ is the dividend rate; which is estimated by the ratio of dividends to the sum of <i>L</i> and V_e (market value of common equity); <i>T</i> is time to maturity for both of call option and liabilities.	The probability of failure is extracted as the probability that call option expires worthless at the end of maturity data - i.e. the value of the company's assets (V_a) be less than the face value of its debt liabilities (<i>L</i>) at the end of the holding period [$P(V_a < L)$]. In Hillegeist et al. (2004), V_a and σ_a are estimated by solving a system of equations; i.e. the call option equation (1) and the optimal hedge equation (2). $\begin{cases} V_e = V_a e^{-\delta T} N(d_1) - L e^{-rT} N(d_2) + (1 - e^{\delta T}) N(d_1) V_a & (1) \\ \sigma_e = \frac{V_a e^{-\delta T} N(d_1) \sigma_a}{V_e} & (2) \end{cases}$ where V_e is the market value of common equity at the time of estimation, σ_e is the annualized standard deviation of daily stock returns over 12 months prior to estimation, <i>r</i> is the risk-free interest rate, and d_1 and d_2 are calculated as follows; $d_1 = \frac{\ln\left(\frac{V_a}{L}\right) + (r - \delta - \frac{1}{2}\sigma_e^2) \times T}{\sigma_e \sqrt{T}}; d_2 = d_1 - \sigma_e \sqrt{T}$ Where $V_{a,t}$ is the value of the company's assets in year <i>t</i> and $V_{a,t-1}$ is the value of the company's assets in year <i>t</i> - 1. Bharath and Shumway (2008) proposed a naïve approach to estimate V_a and σ_a as follows; $V_a = V_e + D; \sigma = \frac{V_e}{V_a} \sigma_e + \frac{D}{V_a} \sigma_d$ where $\sigma_d = 0.05 + 0.25\sigma_e$. Further, the firm's expected return μ is peroxided by the risk-free rate, <i>r</i> or the stock return of previous year restricted to be between <i>r</i> and 100%.
Contingent claim analysis (CAA): Down-and-Out Call (DOC) Barrier Option Model	Jackson and Wood (2013)	A naïve DOC barrier option as an extension of BSM model, which assumes that debt holder's position in the firm is like holding a portfolio of risk-free debt and a DOC option with a strike price (or Barrier) equal to total liabilities (<i>L</i>). The model rests on the assumptions of no dividends, zero rebate, costless failure proceedings, and set return on asset equal to the risk-free rate.

	$P(distress)_i = N \left[\frac{\ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}} \right] + \left(\frac{L}{V_a}\right)^{\frac{2(\mu)}{\sigma_e^2}-1} N \left[\frac{\ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}} \right]$	
Discrete time hazard model (Duration dependent hazard model)	<p>Shumway (2001) $\log [p_{i,t}/(1 - P_{i,t})]$ $= -13.303 - 1.983 NITA + 3.593 TLTA$ $- 0.467 R.size - 1.809 LAGEXRET$ $+ 5.791 SIGMA$ <i>NITA</i>: Net income /Total assets; <i>TLTA</i>: Total liabilities / Total assets; <i>RSize</i>: Relative size; <i>LAGEXRET</i>: Lag of excess return ($r_{it-1} - r_{mt-1}$)</p>	<p>Shumway proposed a discrete time hazard model using an estimation procedure like the one used for estimating the parameters of a multi-period logit model.</p> $P(y_{i,t} = 1 x_{i,t}) = h(t x_{i,t}) = \frac{\exp^{\alpha(t)+X_{i,t}\beta}}{1 + \exp^{\alpha(t)+X_{i,t}\beta}}$ <p>where $h(t x_{i,t})$ represent the individual hazard rate of firm i at time t, $X_{i,t}$ is the vector of covariates of each firm i at time t. Shumway employed a constant time invariant term, say $\ln(age)$, as a proxy of the baseline rate.</p>
Duration-independent hazard model	$h(t x_{i,t}) = h_0 \cdot e^{x_{i,t} \cdot \beta}$ $p(y_{i,t} = 1) = \frac{1}{1 + e^{-x_{i,t} \cdot \beta}}$	<p>where, α_t is the time-varying baseline hazard function related, which could be relate to firm, e.g. $\ln(age)$ or related to macroeconomic variables, e.g. foreign exchange rate.</p>
Duration-dependent hazard model	$h(t x_{i,t}) = h_0(t) \cdot e^{x_{i,t} \cdot \beta}$ $p(y_{i,t} = 1) = \frac{1}{1 + e^{-(\alpha_t + x_{i,t} \cdot \beta)}}$	
Cox hazard framework	$PL(\beta) = \prod_{i=1}^m \left[\frac{\exp(\sum_{j=1}^p \beta_j x_j^i(t))}{\sum_{k \in R_i(t)} \exp(\sum_{j=1}^p \beta_j x_j^k(t))} \right]$ <p>where i is the firm in distress, k is the firm in the risk set at time t, and p is the number of features.</p>	<p>A partial likelihood function on the training sample is used to estimate the coefficients β. This equation estimates β without considering the baseline hazard rate (Hosmer and Lemesho, 1999). However, to use the developed model for estimation of distress probabilities, the baseline hazard rate is required. We follow Chen et al. (2005) in estimating the integrated baseline hazard function with time-varying covariates based on Anderson (1992) as follow:</p> $\hat{H}_0(t) = \sum_{\tilde{T}_i \leq t} \frac{D_i}{\sum_{j \in (\tilde{T}_i)} \exp(\hat{\beta}' \cdot x_j(\tilde{T}_i))}$ <p>where D_i is a dummy variable representing whether firm i faces distress or not, i.e. $D_i = 0$ for non-distress and $D_i = 1$ for distress; \hat{T}_i is the distress time for the ith firm; $\hat{\beta}$ is the vector of estimated coefficients; and \tilde{T}_i is the distress time for the ith firm. Using Equations (4-29) and (4-30), we estimate the probability of distress for individual firms in Equation</p>

